

METHOD AND DEVICE FOR SPEECH PROCESSING

CLAIM FOR PRIORITY

This is a national stage application of PCT/DE00/01116,
5 which was published in the German language on January 11,
2001, which claims the benefit of priority to German
Application No. 19931050.5, filed in the German language
on July 6, 1999.

TECHNICAL FIELD OF THE INVENTION

10 A system and method for speech processing, and in
particular, an orthographic input is converted into a
phonetic transcription the conversion result is checked
and corrected.

BACKGROUND OF THE INVENTION

15 The development of workaday speech recognition systems and
speech control systems has for years been one of the main
lines of development of computer technology. In the course
20 of this development, substantial advances have been
achieved and marketable speech recognition systems have
been established which are also proving themselves in
practical use. Advanced systems of this type are also
fundamentally suited for speech control of a computer
25 and/or of connected peripherals. Simple speech recognition
systems, which can, however, process only a relatively
small vocabulary, are also already in use in the sectors
of consumer electronics and motor vehicle equipment, as
well as further sectors in which acoustic control of
30 equipment on the basis of a limited vocabulary is possible
and sensible.

As a rule, in the case of speech recognition systems there
are tools which can be used to input the vocabulary to be
35 recognized by the speech recognition system. As a rule,



the words or utterances are input in orthographic notation via an appropriate interface software of the computer program and are automatically converted into the internal notation of the speech recognition system (mostly a variant of phonetic transcription (phonetic script)). In this conversion process, which is automatic or supported by lexicon look-up, errors can occur in the phonetic transcription which arise from inadequate conversion rules and/or incomplete lexica. Since the speech recognition system builds up its recognition process on the basis of the phonetic transcription thus generated, an incorrect phonetic transcription also produces errors in the speech recognition.

In order to ensure optimum performance, it must be ensured that the phonetic transcription is as correct as possible.

The problem has so far been solved in that the user has been able to check manually the phonetic transcription generated by the system after inputting of the orthographic (correct) notation. However, this is difficult, as a rule, for untrained staff. Consequently, use has been made of various aids on offer in SW on the market:

1. The user can have displayed for himself words which are typical of the various phonetic symbols and in which such symbols are contained, and can correct the phonetic notation manually. In this case, he is further supported in a few systems to the effect that no incorrect character sequences of the phonetic transcription can be used, since the software employed can input only those character strings which represent a valid ASCII sequence for the phonetic character set used.

2. The phonetic transcription is converted again into an audible speech from the phonetic notation with the aid of text-to-speech software systems, that is to say speech synthesizing methods. This serves the purpose of the acoustic plausibility check of the phoneme string which has been automatically generated by the system for a word. This audible test can, however, eliminate only drastic errors and is subject to the shortcomings of the acoustic channel. Moreover, it is necessary to ensure correspondence between the phonetic alphabets used in the speech recognition similar to the speech synthesis, and this applied to very few cases.

SUMMARY OF THE INVENTION

15 The invention is based on a method and device for speech processing which are designed, in particular, to improve user-friendliness and, in conjunction therewith, also by enhanced accuracy and reliability.

20 The invention includes replacing the outputting of a word converted into phonetic transcription, unfamiliar to, and can be handled only with difficulty by the linguistically untrained user. Typically, these phonetic scripts are phonetic script by an outputting which is simple and can be handled more reliably. The output selected forms a "pseudo-orthographic" and does not demand of the user knowledge of special characters of the phonetic transcription and of their special rules. Put simply, the outputting of the converted words is performed "in the way they are spoken".

This pseudo-orthographic outputting, which is easy to understand even for the layman and can be effectively handled, of a language converted into phonetic transcription requires an additional step in the speech

processing method. Specifically, the step of conversion from the phonetic transcription into this pseudo-orthographic representation. This additional step includes a method which the phonetic units of the words are converted, in a self-learning fashion or with access to a predetermined set of rules, into simple graphemic units of written script. This conversion is performed in a simple and preferred embodiment by accessing a stored phoneme/grapheme assignment table which is initialized at least with an initial stock of assignment rules and can, if appropriate, be extended by the user in the course of a self-learning process during the application of the system on the basis of additional inputs.

In one embodiment, the self-learning process mentioned, the method also comprises a conversion step of reverse conversion into the phonetic transcription from a pseudo-orthographic representation (employed by the user when inputting for the purpose of correcting the primary conversion result). The tabular assignment mentioned can also be used in this step and, if appropriate, can be supplemented and refined in the course of a self-learning process.

One embodiment of the invention includes, in addition to a first converter unit for converting an orthographic input into the phonetic transcription, a second converter unit for converting from the phonetic transcription into the pseudo-orthographic representation mentioned, and an output unit for outputting in this form of representation.

The invention may also include a third converter unit for the abovementioned development of the method, which permits the user to make a correcting input by using the pseudo-orthographic representation.

In order to apply the phoneme/graphème assignment table mentioned, in a preferred embodiment the device has an appropriate memory in which this assignment table is held
5 accessibly for the second and/or third converter unit.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a speech processing device according to the invention.

10

DETAILED DESCRIPTION OF THE INVENTION

The figure shows a schematic illustration of a speech processing device 1 for carrying out the method according to the invention in an embodiment in the form of a
15 functional block diagram. The speech processing device 1 comprises an acoustic input unit 3 at whose output a preprocessed stream of speech S1 is present which is fed to an input of a speech recognition unit 5 which outputs a written text S2. The speech recognition unit 5 comprises a
20 vocabulary memory 5a in which the vocabulary of the speech recognition unit is stored in the phonetic notation customary in conventional speech recognition systems.

The vocabulary memory 5a is continuously updated by the
25 input of additional terms by means of an alphanumeric input unit 7, which terms are converted from the orthographic input format in a first converter unit 9 into the phonetic transcription (phonetic script). A lexicon memory 11 supports the conversion procedure in the first
30 converter unit 9. For the purpose of checking and correcting undertaken inputs, a second converter unit 13 is provided for converting the phonetic transcription into a pseudo-orthographic representation. This is indicated on a display screen 15 for the user.

35

Also provided is a third converter unit 17 for converting pseudo-orthographic inputs via the alphanumeric input unit 7 into phonetic notation, the output of which is connected to the vocabulary memory 5a of the speech recognition unit 5. The second and third converter units 13, 17 are assigned an assignment memory 19, organized in the form of a look-up table, for predetermined phoneme/grapheme assignments.

10 An input, performed by the user, of a new term in correct orthographic notation is converted in the first converter unit 9 into phonetic script and can - depending on the actual organization of the system - already be fed in this form to the vocabulary memory 5a. In each case, the word
15 converted into phonetic script is fed, however, to the second converter unit 13, where a further conversion into a pseudo-orthographic representation is performed, which is displayed on the display screen 15 and causes the user, if appropriate via the input unit 7 - now in the pseudo-
20 orthographic representation, which also appears on the display screen - to make a correcting input, or else to confirm the displayed pseudo-orthographic representation. The pseudo-orthographic input is converted in the third
25 converter unit 17 into phonetic script and now (for the first time or, if the word has already been taken over into the vocabulary memory 5a on the occasion of the first input, in a correction mode) fed to the vocabulary memory 5a. The contents thereof are thereby expanded by a word checked with regard to the phonetic notation.

30 The procedure described above is explained below using two examples:

1st example

35

"Jacques Chirac" is input in correct orthographic notation via the alphanumeric input unit 7. The phonetic notation "sh a xk sh i: rr a xk" is formed therefrom in the first converter unit 9. The second converter unit 13 forms "sch
 5 a k sch i r a k" therefrom, and the input name is displayed on the display screen 15 in this notation. It is possible - without knowing the phonetic alphabet used in the first conversion - to perceive from this representation that the phonetic notation generated by the
 10 system is adequate. The user can confirm the conversion result, and the newly input name passes (in phonetic notation) into the vocabulary memory 5a.

2nd example

15 "Professional Service" is input via the input unit 7. The first converter unit 9 generates therefrom in phonetic notation
 "p r o: f ae sh o n :e: ll s oe r v i: cc :e". In the result of the further conversion in the second converter
 20 unit 13, "Profäschonell Sörwieke" is yielded therefrom in pseudo-orthographic notation, and this representation is again displayed on the display screen 15.

The user perceives straight away that the phonetic script
 25 generated by the system cannot be correct, since it does not correspond to the usual pronunciation of the input word combination. The user will now use the input unit in conjunction with the pseudo-orthographic notation, which is illustrated on the screen, to undertake a correction,
 30 and the correction result is converted again in the third converter unit 17 from the pseudo-orthographic notation into the phonetic one, and taken over in this form into the vocabulary memory 5a. In the example given, the user will therefore input "Profäschonnell Sörwis", and the new

word combination (in phonetic notation) is anchored in the vocabulary memory.

The method can also be carried out in a plurality of steps
5 when, after a first correction by the user, a further
conversion from the phonetic notation into the pseudo-
orthographic one is performed in conjunction with a
further display in this representation such that, if
appropriate, system errors can be eliminated iteratively.
10 In this case, it is preferred to apply a self-learning
system for example in the form of a neural network with
the aid of which a self-adaptation of the memory contents
of the assignment memory 19 and/or the assignment rules of
the first conversion operation (orthographic - phonetic)
15 can be performed.

The design of the invention is not limited to the example
described above, but is also possible in a multiplicity of
modifications which are within the scope of expert
20 activity.